



UNITED STATES PATENT AND TRADEMARK OFFICE

24

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/033,199	12/28/2001	Shivnath Babu	BABU 1-10-42	8231
47394	7590	06/03/2005	EXAMINER	
HITT GAINES, PC LUCENT TECHNOLOGIES INC. PO BOX 832570 RICHARDSON, TX 75083				LERNER, MARTIN
		ART UNIT		PAPER NUMBER
		2654		

DATE MAILED: 06/03/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	10/033,199	BABU ET AL.
	Examiner	Art Unit
	Martin Lerner	2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 07 March 2005.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1 to 24 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1 to 24 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 07 March 2005 is/are: a) accepted or b) objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) <input type="checkbox"/> Notice of References Cited (PTO-892)	4) <input type="checkbox"/> Interview Summary (PTO-413)
2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)	Paper No(s)/Mail Date. _____
3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)	5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)
Paper No(s)/Mail Date _____	6) <input type="checkbox"/> Other: _____

DETAILED ACTION

Specification

1. The disclosure is objected to because of the following informalities:

On page 56, ¶ [0130], "states" should be —stated—.

On page 57, ¶ [0133], the comma should be removed between "different" and "CaRTs".

On page 58, ¶ [0135], "decreased" should be —decrease—.

Appropriate correction is required.

Claim Rejections - 35 USC § 102

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless —

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. Claims 1, 9, and 17 are rejected under 35 U.S.C. 102(e) as being anticipated by *Fayyad et al. ('882)*.

Fayyad et al. ('882) discloses a system and method for database management, comprising:

"a table modeler that discovers data mining models with guaranteed error bounds of at least one attribute in said data table in terms of other attributes in said data table" – the invention evaluates a data database 10 having many records stored on storage devices; each record in the database 10 has many attributes or fields which for a representative database might include age, income, number of years of employment, census data, etc. (column 4, line 60 to column 5, line 2); implicitly, a plurality of records, where each record has a number of attributes is a table; a data clustering model ("table modeler") is produced that implements a data mining engine for answering queries about data records in the database (column 5, lines 20 to 25); accuracy parameters ("guaranteed error bounds") are used to control the clustering; an accuracy parameter can be the percentage by which the number of points is allowed to deviate from an expected value or the probability of a tile satisfying the accuracy criterion (column 9, line 63 to column 10, line 42);

"a model selector, associated with said table modeler, that selects a subset of said at least one model to form a basis upon which to compress said data table" – a data mining engine 12 forms conclusions about the accuracy of an initial model (M), and the model is refined until the model more accurately represents the data stored in the database (column 9, lines 37 to 62); a cluster must satisfy an accuracy requirement for the model to be judged suitable (column 10, lines 33 to 42); a model represents a compressed version of records in data database 10 (Abstract).

Claim Rejections - 35 USC § 103

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. Claims 2, 4, 5, 8, 10, 12, 13, 16, 18, 20, 21, and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal* ('311).

Concerning claims 2, 10, and 18, *Fayyad et al.* ('882) does not disclose specifics about the modeling process as employing classification and regression tree (CaRT) data mining to model attributes. However, *Agrawal* ('311) suggests data mining with decision trees for modeling records having one or more attribute values may be by classification and regression trees. (Column 5, Line 63 to Column 6, Line 7; Column 6, Lines 58 to 67) The stated objective is provide an efficient method for generating a decision-tree classifier that is compact, accurate, has short training times and is scalable. (Column 3, Lines 11 to 24) It would have been obvious to one having ordinary skill in the art to employ classification and regression trees for data mining of model attributes as taught by *Agrawal* ('311) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of generating decision trees by a classifier that is compact, accurate, has short training times and is scalable.

Concerning claims 4, 12, and 20, *Agrawal* ('311) discloses pruning for short training time (column 8, line 40 ff).

Concerning claims 5, 13, and 21, *Agrawal* ('311) discloses pruning for representing misclassification errors based upon encoding costs (column 9, lines 34 to 54), which is equivalent to a "scoring-based method".

Concerning claims 8, 16, and 24, *Agrawal* ('311) discloses a greedy algorithm may be used for subsetting (column 8, line 3).

6. Claims 2, 3, 10, 11, 18, and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Pednault*.

Fayyad et al. ('882) does not disclose specifics about the modeling process as employing classification and regression trees or a Bayesian network. However, *Pednault* teaches a method for constructing predictive models that involve Bayesian networks (column 2, lines 20 to 30 and column 2, lines 45 to 52) and classification and regression trees (column 2, lines 35 to 45). The objective is to provide a method of handling missing values. It would have been obvious to one having ordinary skill in the art to employ classification and regression trees or Bayesian networks as suggested by *Pednault* in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of providing a method for handling missing values.

7. Claims 6, 14, and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Chakrabarti et al.* ('005).

Fayyad et al. ('882) omits selecting a subset based upon a compression ratio. However, *Chakrabarti et al.* ('005) teaches a method for data mining where a

compression ratio is an indicator of complexity of compressed files. (Column 16, Lines 18 to 25) The objective is to select candidate data patterns from a dataset based on the variations of support values of a pattern. (Column 5, Lines 4 to 14) It would have been obvious to one having ordinary skill in the art to select a data subset based upon a compression ratio as suggested by *Chakrabarti et al.* ('005) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of selecting candidate data patterns from a dataset.

8. Claims 7, 15, and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal et al.* ('048).

Fayyad et al. ('882) does not disclose that a process by which a model selector selects a subset is NP-hard. However, *Agrawal et al.* ('048) teaches that, in general, an optimized rule mining problem is NP-hard. (Column 4, Lines 9 to 14) The objective is to provide a method for identifying database association rules which are optimal at upper and lower support-confidence borders. (Column 4, Line 30 to Column 5, Line 45) It would have been obvious to one having ordinary skill in the art that model selection is an NP-hard algorithm as suggested by *Agrawal et al.* ('048) in the multi-dimensional database record compression of *Fayyad et al.* ('882) for the purpose of providing optimal association rules at upper and lower support-confidence borders.

Response to Arguments

9. Applicants' arguments filed 07 March 2005 have been fully considered but they are not persuasive.

Applicants argue that *Fayyad et al.* ('882) does not teach compression of a data table including at least one attribute in the data table in terms of other attributes in the data table and selecting a subset of the at least one model to form a basis upon which to compress the table. Applicants maintain that they do not find where *Fayyad et al.* ('882) teaches selecting a subset of a data mining model for any purpose. This position is traversed.

Fayyad et al. ('882) discloses a compression scheme to characterize a database containing data records. (Abstract) The compression scheme represents data records in a database by a model of the data records rather than the data records *per se*. Thus, the database is made to be more compact, and storage space is saved, if actual data is not used in answering queries. (Column 5, Lines 5 to 10) The database of records includes attributes of age, income, number of years of employment, vested pension benefits, etc. (Column 4, Line 60 to Column 5, Line 2; Column 5, Lines 25 to 39; Table 1) Querying a database of records on the basis of attributes including age, income, years of employment, vested pension benefits, etc., to discover information about the records in the database is known as "data mining". Thus, *Fayyad et al.* ('882) clearly discloses compressing a data table including at least one attribute for purposes of data mining using at least one model.

Moreover, *Fayyad et al.* ('882) discloses selecting a subset of a model in order to compress the information in the database of data records. For *Fayyad et al.* ('882), the models are called clustering models. The objective is to find a set of data points that best represent a cluster, and then to represent the properties of the cluster from those data points. (Column 5, Line 45 to Column 6, Line 6) Heuristically, each cluster is represented by less than all the set of data records in that cluster. That is, a subset of all the records for a cluster represent that cluster, so if each cluster is represented by a subset of records, then all the clusters must be represented by a subset of records. In fact, however, each cluster is represented by a compact representation of multivariate Gaussian mixture models containing a sufficient number of components, where those skilled in the art recognize that a Gaussian mixture model is a statistical representation of a plurality of multidimensional data points.

Both a K-means clustering process and an Expectation-Maximization (E-M) clustering process are disclosed. (Column 6, Lines 7 to 32) K-means clustering simply assigns each point to one of the clusters, so that each cluster represents a subset of the data for that cluster, and the cluster model is a sum of all clusters, also a subset of the data. E-M clustering applies a weighting factor to each point in a cluster.. (Column 6, Lines 33 to 41) In either case, each cluster in the model is represented or summarized as a multivariate Gaussian have a probability distribution function of n-dimensional data points $x = (x_1, \dots, x_n)$, where each dimension represents one of the attributes of a data record. (Column 7, Lines 20 to 34)

The question is to determine the number of clusters K, or partitions of data, in a model, so as to best represent a model. Clearly, a greater number of clusters provides a more accurate model, but a greater number of clusters also increases costs in terms of time and memory to access and store the model. Thus, the clustering procedure is designed to iteratively determine an optimal number of clusters until a stopping point is reached. (Column 7, Line 66 to Column 8, Line 9) A starting point of K clusters is first selected, and then the number of clusters is increased, or grown, so that the new model better fits the data and improves accuracy. (Column 8, Lines 10 to 24) However, a model comprising all of the clusters is still a subset of all the data records in the database. It follows that a clustering procedure that selects a desired number of clusters involves selecting a subset of a data mining model.

Fayyad et al. ('882) states that an accuracy parameter is used to control the clustering process. (Column 9, Line 63 to Column 10, Line 32) The accuracy parameter corresponds to Applicants' "guaranteed error bounds" between attributes. An accuracy parameter describes the percentage by which the number of points is allowed to deviate from an expected value. If a percentage of points exceed an accuracy criterion, then this indicates that the number of clusters K should be increased. The accuracy criteria determine a granularity of the clusters as being low or high. A lower granularity of a cluster indicates there are more points per cluster, and a higher granularity of a cluster indicates there are fewer points per cluster. Thus, the accuracy parameters provide "error bounds" for a model, so that if further accuracy is desired ("guaranteed"), then the number of clusters must be increased.

Applicants' contention that *Fayyad et al.* ('882) omits selecting a subset of a data mining model is incorrect. *Fayyad et al.* ('882) discloses selecting a subset of at least one data mining model during the clustering process. Each cluster model represents a subset of all the data from a data source. All of the data in a data source represents one model, and any clustering model represents a subset of all the data from a data source. Thus, *Fayyad et al.* ('882) selects a subset of a data mining model during the clustering process.

Therefore, the rejections of claims 1, 9, and 17 under 35 U.S.C. §102(e) as being anticipated by *Fayyad et al.* ('882), of claims 2, 4, 5, 8, 10, 12, 13, 16, 18, 20, 21, and 24 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal* ('311), of claims 2, 3, 10, 11, 18, and 19 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Pednault*, of claims 6, 14, and 22 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Chakrabarti et al.* ('005), and of claims 7, 15, and 23 under 35 U.S.C. §103(a) as being unpatentable over *Fayyad et al.* ('882) in view of *Agrawal et al.* ('048), are proper.

Conclusion

10. **THIS ACTION IS MADE FINAL.** Applicants are reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not

Art Unit: 2654

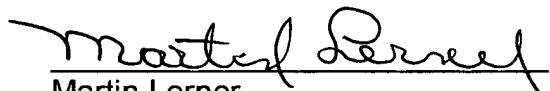
mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

ML
5/25/05


Martin Lerner
Martin Lerner
Examiner
Group Art Unit 2654